論文 94-19-6-16

프랙탈 차원을 이용한 모음인식

正會員 崔 哲 榮* 正會員 金 炯 淳* 正會員 金 在 浩* 正會員 孫 慶 植*

Vowel Recognition Using the Fractal Dimension

Chul Young Choi*, Hyung Soon Kim*, Jae Ho Kim*,
Kyung Sik Son* *Regular Members*

요 약

본 논문에서는 음성신호의 프랙탈 차원을 이용하여 한국어 모음인식 실험을 수행하였다. 프랙탈 차원은 Minkowski-Boulignad 차원을 사용하였으며, 형태학적 커버링(morphological covering) 방법을 이용하여 구하였다. 프랙탈 차원과 더불어 기존에 우수한 음성 인식 파라메타로 알려져 있는 LPC 켑스트럼(cepstrum)을 함께 사용하여 인식실험을 하였으며, 프랙탈 차원의 음성인식에의 유용성 여부를 조사하였다. 다양한 자음환경에서의 모음인식 실험결과, LPC 켑스트럼 만을 사용하는 경우 및 프랙탈 차원과 LPC 켑스트럼을 함께 사용하는 경우의 모음 오인식율이 각각 5.6% 및 3.2%로 얻어졌다. 이는 LPC 켑스트럼에 프랙탈 차원을 추가함으로써 오인식되는 데이타가 40% 이상 감소되는 결과이며, 프랙탈 차원이 음성인식에 있어서 유용한 특징 파라메터임을 보여준다.

ABSTRACT

In this paper, we carried out some experiments on the Korean vowel recognition using the fractal dimension of the speech signals. We chose the Minkowski-Bouligand dimenison as the fractal dimension, and computed it using the morphological covering method. For our experiments, we used both the fractal dimension and the LPC cepstrum which is conventionally known to be one of the best parameters for speech recognition, and examined the usefulness of the fractal dimension. From the vowel recognition experiments under various consonant contexts, we achieved the vowel recognition error rates of 5.6% and 3.2% for the case with only LPC cepstrum and that with both LPC cepstrum and the fractal dimension, respectively. The results indicate that the incorporation of the fractal dimension with LPC cepstrum gives more than 40% reduction in recognition errors, and indicates that the fractal dimension is a useful feature parameter for speech recognition.

*釜山大學校 電子工學科
Dept. of Electronic Engineering, Pusan Univ.
論文番號 : 9437
接受日字 : 1994年 2月 4日

Ⅰ. 서 론

프랙탈 기하학은 아주 불규칙하고 복잡하게 보이는 자연현상 내에서도 그속의 어떤 법칙과 규칙성을 발견하여 그러한 자연현상을 해석하기 위한 학문분야 중의 하나로서, 1970년대 중반 Mandelbrot[1]에 의해 제창되고 이후 Barnsley[2]등 다른 여러 사람에 의해 발전되어왔다. 프랙탈은 부분의 모습이 전체와 닮은 모습을 가지는 형상으로서 구름, 산, 나무, 해안선 등이 대표적인 예이며, 그 외에도 주식시세 변동, 인구증가 변동, 음성 신호 및 일부의 잡음 신호 등의 시계열(time series) 데이타들이 프랙탈로 해석될 수 있다. 그리고 프랙탈의 닮은 정도를 나타내는 프랙탈 차원은 프랙탈을 해석하는데 있어서 매우 중요한 역할을 한다. 닮은 정도가 큰 프랙탈일수록 높은 차원을 가지며 보다 거칠고 불규칙하게 보인다.

근래에 들어 음성신호처리 분야에서도 음성의 프랙탈 특성을 이용하려는 연구가 점차적으로 이루어지고 있다.[3][4] 음성의 발생 과정에는 공기 흐름이 비선형적인 역학성으로 인한 어떤 크고 작은 정도의 혼란이 존재한다.[3] 이러한 특성이 음성신호에 반영되어 음성신호의 여러 배율(magnification)의 크기에서 서로 유사한 모습이 나타난다. 따라서, 음성신호는 프랙탈로 해석될 수 있으며 프랙탈 차원을 새로운 파라메타로서 음성인식을 비롯한 음성신호처리 분야에 이용할 수 있다.

이를 바탕으로 본 논문에서는 음성인식에 프랙탈 차원을 이용하였다. 기존에 가장 우수한 음성특징 파라메타 중 하나로 알려져 있는 LPC 켑스트럼(cepstrum)에 프랙탈 차원을 추기로 사용하여 VQ(Vector Quantization)를 이용한 화자종속 모음인식 실험을 하였다. 실험결과 인식율이 개선됨을 확인하였으며, 이로써 프랙탈 차원이 음성인식에 있어서 하나의 훌륭한 파라메터가 될 수 있음을 보였다. 본 논문에서는 음성신호의 프랙탈 차원을 구하는 방법으로서 다른 프랙탈 차원보다 구현이 간단하고 강인한 특성을 지니는 것으로 알려진 Minkowski-Bouligand 차원을 사용하고 이를 형태학적 커버링(Morphological Covering) 방법을 이용하여 구하였다.[5]

본 논문의 구성은 다음과 같다. 서론에 이어 2장에서는 음성신호의 프랙탈 차원을 계산하는 방법을 설명하고 3장에서는 프랙탈 차원을 이용한 모음인식 실험에 대해 설명한다. 그리고 4장에서는 인식 실험결과를 검토하여 프랙탈 차원의 음성인식에 있어서의 유용성을 조사하고, 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 음성신호의 프랙탈 차원 계산

음성신호의 프랙탈 차원을 구하기에 앞서, 본 장에서는 먼저 프랙탈 차원의 일반적인 개념과 계산이 간단하면서도 강인한 특성을 갖는 것으로 알려진 Minkowski-Bouligand 차원에 대해 살펴본다. 그 다음으로 이산신호 형태의 음성신호의 프랙탈 차원을 구하기 위한 방법으로서 Maragos와 Sun이 제안한 형태학적 커버링 방법에 의해 Minkowski-Bouligand 차원을 계산하는 과정을 살펴본다.[5]

1. 프랙탈 차원의 개념 및 Minkowski-Bouligand 차원

평면상의 한 프랙탈 곡선 X에 대한 프랙탈 차원의 개념은 다음과 같이 설명될 수 있다. 프랙탈은 부분이 전체와 닮은 모습을 가지고 있는 것이므로 X를 확대하여 보았을때 확대하기 전의 X와 닮은 모습을 지니고 있을 것이다. 이때 X의 길이를 구하고자 한다. 길이가 ε인 측정자를 가지고 곡선 X를 따라가면서 놓아갈때 X를 전부 다 덮을 수 있는 측정자의 갯수를 $N(\varepsilon)$개라고 하자. 그러면 이 측정자에 의한 X의 길이 $L(\varepsilon)$는 $N(\varepsilon)\times\varepsilon$로 주어진다. 이때 측정자의 길이 ε을 점차로 작게 하여 나가보자(이는 X를 확대하여 보는것과 같다). 그러면 X의 아주 미세한 부분까지 측정됨으로 인하여 ε의 감소하는 비율보다 $N(\varepsilon)$의 증가하는 비율이 더 크게 되어 측정 되어지는 X의 길이는 점점 늘어나게 된다. 이때 프랙탈 곡선 X의 길이 $L(\varepsilon)$은 ε이 영으로 수렴함에 따라 다음 수식으로 근사화 될 수 있다.

$$L(\varepsilon) \approx (\text{상수}) \times \varepsilon^{1-D},\ \varepsilon \to 0. \qquad (1)$$

여기서 상수 D를 X의 프랙탈 차원이라 부르며 D가 크면 클수록 ε의 감소에 대한 $L(\varepsilon)$의 증가 비율이 커진다. 일반적으로 잘 알려져 있는 프랙탈 차원들은 모두 (1)식에 기초를 두고 있으며 (1)식에서 길이를 어떤 방법으로 구하느냐에 따라 서로 조금씩 다른 특징을 가지고 있다. 프랙탈 차원은 Hausdorff-Besicovitch차원, Minkowski-Bouligand 차원, Box Counting 차원, Entropy 차원 등의 여러가지가 있으나, 그 중에서 Minkowski-Bouligand 차원이 다른 방법들에

비해 구현이 간단하고 강인한 특성을 가지는 것으로 알려져 있으며[5] 이하에 Minkowski-Bouligand 차원에 대해 보다 상세하게 설명한다.

Minkowski-Bouligand 차원의 경우, 앞서 설명한 프랙탈 곡선 X의 길이를 측정하기 위해 X의 Minkowski 덮개(cover)를 사용한다. X의 Minkowski 덮개는 X를 반경 ε의 원으로서 형태학(morphology)에서의 팽창(dilation)을 한것과 같다. 그리고 Minkowski 덮개 면적을 $A(\varepsilon)$라고 하고 X의 길이는 $\lim_{\varepsilon\to 0} L(\varepsilon)$로 주어진다. 여기서 $L(\varepsilon)$는 반경 ε인 원을 이용한 Minkowski 덮개를 사용했을때의 X의 길이의 측정치로서 $L(\varepsilon)$는 $L(\varepsilon)=A(\varepsilon)/2\varepsilon$라는 관계가 성립한다. 이때 (1) 식의 개념으로부터 Minkowski-Bouligand 차원 D_M은 다음과 같이 정의된다.

$$D_M=\lim_{\varepsilon\to 0}\left(\frac{\log[A(\varepsilon)/\varepsilon^2]}{\log 1/\varepsilon}\right) \tag{2}$$

2. 이산신호의 형태학적 커버링 방법

형태학적 커버링 방법은 Minkowski-Bouligand 차원 D_M을 구하는 과정에서 원을 이용한 Minkowski 덮개 대신에 다른 임의의 형태를 가지는 평면내 집합을 이용한 형태학적 덮개(morphological cover)를 사용한 것으로 연속신호 뿐만 아니라 이산신호에 대해서도 효과적으로 적용될 수 있다.

f[n], n=0, 1,…,N을 유한한 길이의 이산신호라고 하자. 이때 (n, f[n])은 Cartesian 좌표의 (x, y)에 해당한다. 그리고 B를 볼록(convex)하고, x, y축에 대해 대칭적인 평면내의 이산집합이라고 하자. $\varepsilon B=\{\varepsilon b : b\in B\}$이고 ε은 1, 2…의 정수이다. 이때 형태학적 팽창 연산자인 $\oplus$을 사용한 f[n]의 형태학적 덮개 $C_B[\varepsilon]$는 다음과 같이 정의 된다.

$$C_B[\varepsilon]\equiv f[n]\oplus\varepsilon B \tag{3}$$

여기서 만약 B가 원이라면 $C_B[\varepsilon]$는 앞서 설명한 Minkowski 덮개와 같다.

다음으로 $C_B[\varepsilon]$의 포락선을 이용하여 $C_B[\varepsilon]$의 면적인 $A[\varepsilon]$를 구한다. g[n]을 B의 상위 포락선이라 하고 $g_\varepsilon[n]$을 εB의 상위 포락선이라 할때 $C_B[\varepsilon]$의 상위 포락선은

$$f\oplus g_\varepsilon=f\oplus g^{(\oplus\varepsilon)}=((f\oplus g)\oplus g\cdots)\oplus g \tag{4}$$

이며 하위 포락선은

$$f\ominus g_\varepsilon=f\ominus g^{(\ominus\varepsilon)}=\underbrace{((f\ominus g)\ominus g\cdots)\ominus g}_{\varepsilon\text{번}} \tag{5}$$

이다. 그리고 일반적으로 B가 크면 형태학적 덮개가 f[n]의 파형의 변화 특성을 잘 나타내지 못하게 되므로, B는 (x, y) 이산좌표상에서 3×3 크기의 점들로 구성된 집합의 부분집합을 갖도록 권장된다.[5] 이때 식 (4)와 (5)는 구체적으로 다음과 같이 표현될 수 있다.

$$f\oplus g[n]=\max_{-1\le i\le 1}\{f[n+i]+g[i]\},\quad \varepsilon=1$$

$$f\oplus g^{(\oplus\varepsilon)}[n]=(f\oplus g^{(\oplus(\varepsilon-1))})\oplus g[n],\quad \varepsilon\ge 2. \tag{6}$$

$$f\ominus g[n]=\max_{-1\le i\le 1}\{f[n+i]-g[i]\},\quad \varepsilon=1$$

$$f\ominus g^{(\ominus\varepsilon)}[n]=(f\ominus g^{(\ominus(\varepsilon-1))})\ominus g[n],\quad \varepsilon\ge 2. \tag{7}$$

이들 상위 포락선과 하위 포락선을 이용하여 $C_B[\varepsilon]$의 면적 $A[\varepsilon]$은 다음과 같이 구해질 수 있다.

$$A[\varepsilon]=\sum_{n=0}^{N}((f\oplus g^{(\oplus\varepsilon)})-(f\ominus g^{(\ominus\varepsilon)}))[n] \tag{8}$$

그리고 마지막으로 (8)에서 구한 $A[\varepsilon]$를 식 (2)에 적용함으로써 D_M을 구한다. 이상의 방법으로 프랙탈 차원 D_M을 구하는 것을 형태학적 커버링 방법이라고 한다.

본 논문에서는 식 (6), (7)에서 g[n]=0 (n=−1, 0, 1)인 수평선분 B를 사용했는데, 이 경우 f[n]의 어떠한 이동이나 크기 변화에 대해서도 f[n]의 프랙탈 차원은 변하지 않는다는 장점을 가진다.[5] 즉, 상수 a 및 b에 대해 f'[n]=a·f[n−n_0]+b이라고 할때 $D_M(f)=D_M(f')$이다.

이러한 장점으로 인해 본 논문에서 음성신호의 프랙탈 차원을 구하는데 있어서 형태학적 커버링 방법에 의한 Minkowski-Bouligand 차원을 사용하였다. 참고적으로 ε이 10, 30일때 g[n]=0 (n=−1, 0, 1)을 사용하여 음성신호의 상위 포락선과 하위 포락선을 구한 예를 그림 1에 나타내었다. 일반적으로 프랙탈 차원을 구하는 방법으로 잘 알려진 Box Counting 차원은 위의 f[n]과 교차하는 사각형 수와 f'[n]과 교차하는 사각형 수가 서로 달라 이로 인하여 프랙탈 차원이 달라지는 문제점이 있어 음성신호의 프랙탈 차원을 구하는데 적합하지 않다.[3][5]

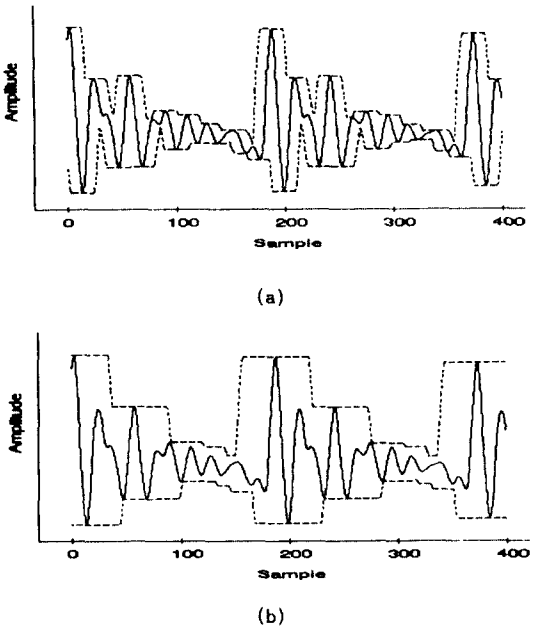

그림 1. 음성 신호 '아'(실선)에 대한 포락선(점선)의 예. (a) $\varepsilon=10$. (b) $\varepsilon=30$.

Fig 1. Example of envelopes(dashed line) for the speech signal 'a'(solid line). (a) $\varepsilon=10$. (b) $\varepsilon=30$.

그리고 프랙탈 차원 D_M을 구하는 과정에서 마지막으로 한가지 고려해야 될 사항이 있는데, 이는 위에서 (8)식을 (2)식에 적용할때 ε을 무한히 0으로 보낼 수 없다는 것이다. 그러므로 실제로는 어떤 임의의 ε 범위에서 $\log(A[\varepsilon]/\varepsilon^2)$ 대 $\log(1/\varepsilon)$의 그래프를 그리고, 최소 자승법을 사용하여 그래프를 직선으로 맞춘 다음 그 직선의 기울기를 프랙탈 차원의 근사적인 값으로서 사용한다. 그런데 음성신호와 같은 실 세계(real world)의 프랙탈은 수학적으로 생성된 Koch 곡선[6]과 같은 이상적인 프랙탈 곡선과는 달리 모든 크기의 배율에 걸쳐 같은 크기의 프랙탈 차원을 가지지 않는다.[3][5] 즉 모든 크기의 배율에 걸쳐 동일하게 닮은 구조를 가지지 않는다는 것이다. 그러므로 어떤 크기의 배율에서 보느냐에 따라 프랙탈 차원의 측정치가 달라지는데 이것은 어떤 범위의 ε값을 사용할것인가 하는 문제를 발생시킨다.

ε의 범위에 대해서 논의할때 두가지 변수가 존재하는데 첫째는 어떤 ε의 값에서 시작하는가 이며, 둘째는 ε 범위의 크기를 얼마로 잡는가 하는 것이다. 이 두가지 문제는 임의이 ε 범위를 $\{\varepsilon_0, \varepsilon_0+1, \varepsilon_0+2, \cdots, \varepsilon_0+m-1\}$이라고 놓았을때 어떤 ε_0과 m의 값을 선택할 것인가 하는 문제와 같다. ε_0의 경우 ε_0의 값이 크면 형태학적 덮개를 사용하는 과정에서 지나친 smoothing이 이루어져서 신호파형의 변화특성을 잘 살리지 못하는 경향이 있게 되므로, ε_0의 값을 작게 하는 것이 좋다. 그리고 m값의 결정은 경험적인 문제로서 이전의 많은 연구자들이 경험적으로 결정해왔다. 앞의 제한된 ε의 범위에서 구한 프랙탈 차원을 국부적 프랙탈 차원(Local Fractal Dimension, LFD)[3]이라고 하며 본 논문에서는 m의 값에 따라 mLFD이라고 이름을 붙였다. 본 논문의 음성인식 실험에서는 ε_0를 1로 둔 10, 20, 30 세 값의 m을 사용하였으며 각각에 대해 인식 결과를 비교하였다. 참고적으로 그림 1의 음성신호에 대해 ε_0이 1이고 m이 20일때에 대해 line fitting 한 예를 그림 2에 나타내었다. 그리고 역시 그림 1의 음성신호에 대해 ε_0과 m의 변화에 따른 프랙탈 차원의 변화를 그림 3에 나타내었다. 이때 음성신호에 따라 ε_0에 대한 프랙탈 차원의 변화특성이 서로 다른데 이를 이용한 음성신호의 분류도 연구되고 있다.[3]

프랙탈 차원을 구하는데 소요되는 계산량은 한 frame에 포함되는 음성신호 샘플 수 N과 LFD의 m값에 따라 다음과 같이 계산될 수 있다. 먼저 Minkowski 덮개에 의한 m개의 면적 $A[\varepsilon]$, $\varepsilon=1,\cdots,m$,은 식 (8)과 같이 주어지며, 이에 따른 계산량은 4mN회의 integer 비교 및 m(2N+1)회의 integer 덧셈/뺄셈이다. 그리고 이들 $A[\varepsilon]$, $\varepsilon=1,\cdots,m$, 값들을 이용해서 $\log(A[\varepsilon]/\varepsilon^2)$ 대 $\log(1/\varepsilon)$ 그래프의 기울기를 최소자승법에 의해 구하는 과정은 일반적인 선형회기분석의 경우와 같으며, 2m회의 log연산(그중 m회는 미리 계산 가능하므로로 제외될 수 있음)과 더불어 7m+5회의 부동소수점 곱셈, 그리고 4m+2회의 부동소수점 덧셈으로 구성된다. 실제로 m값에 비해 N값이 상당히 크기 때문에(N=200, m=10, 20, 또는 30), 대부분의 계산이 $A[\varepsilon]$를 구하는 과정에 소요된다. C 언어에 의한 프로그램 결과에 따르면 m=10일 때의 프랙탈 차원의 계산량은 12차 LPC 켑스트럼의 계산량과 거의 비슷하며, m=20 또는 30일 경우 이들의 2배 및 3배의 계산량이 소요되게 된다. 그러나 일반적으로 음성인식에서 음성특징추출에 소요되는 계산량은 전체 계산량의 상당히 작은 부분에 불과함을 고려할 때 프랙탈 차원의 도입에 따른 계산량 증가는 큰 문제가 되지 않는다고 판단된다.

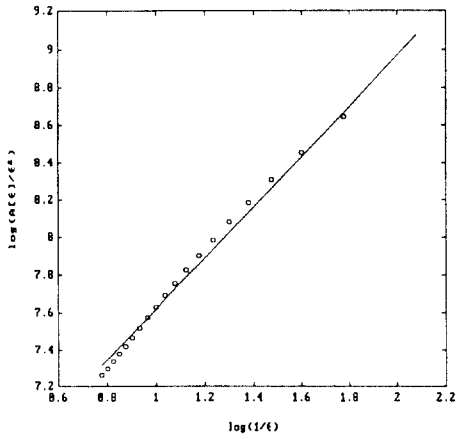

그림 2. 직선 맞춤에 의해 프랙탈 차원을 구하는 예 ($\varepsilon_0 = 1$, m = 20인 경우).

Fig 2. Example of calculating the fractal dimension by line fitting (In case of $\varepsilon_0 = 1$ and m = 20)

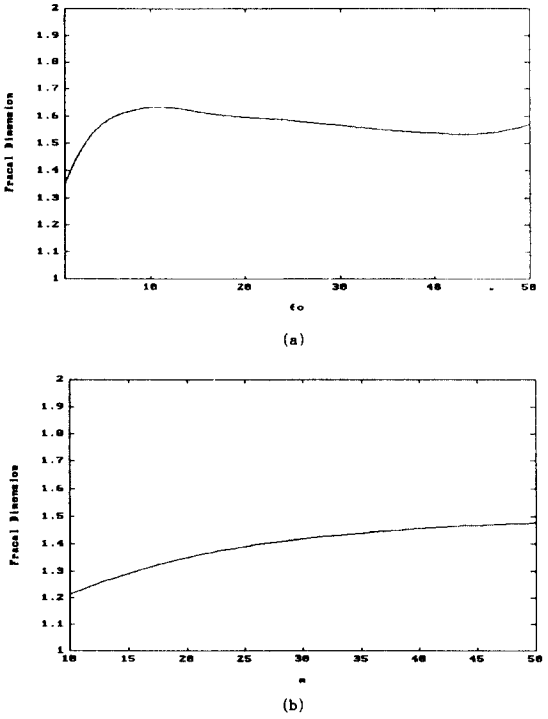

그림 3. ε_0 및 m의 변화에 따른 프랙탈 차원의 변화.
(a) m = 20으로 고정하고 ε_0를 변화시킨 경우.
(b) $\varepsilon_0 = 1$로 고정하고 m을 변화시킨 경우.

Fig 3. Variation of fractal dimension according to the variatio of ε_0 and m.
(a) In case of variable ε_0 and fixed m(m = 20).
(b) In case of variable m and fixed ε_0($\varepsilon_0 = 1$).

Ⅲ. 프랙탈 차원을 이용한 모음인식

이 장에서는 앞의 방법에 의해 구해진 음성신호의 프랙탈 차원을 초성자음과 모음으로 구성된 한국어 CV(Consonant-Vowel) 단음절에서의 모음인식에 적용하였다. 이를 위하여 일반적으로 음성인식에 있어서 가장 효과적인 음성 특징 파라메타의 하나로 알려진 LPC 켑스트럼과 프랙탈 차원 각각에 대해 벡타 양자화(vector quantization)를 이용한 인식 음성인식 실험을 하고 이들을 같이 사용 하였을때 어느 정도 인식율이 증가 하는지를 조사하여 프랙탈 차원의 음성인식에 있어서의 유용성을 조사한다.

1. 실험에 사용한 음성 데이타

한국전자통신연구소(ETRI)에서 구성한 611 고립단어 음성 데이타[7] 중 두 사람의 남자 음성 데이타로서, 한 사람당 126개의 CV 단음절(18개의 자음 × 7개의 모음)을 두 번씩 발음한 것을 사용하였다. 하나의 모음당 18개의 선행 자음이 있고 각각의 음성 데이타는 하나의 자음과 모음으로서 이루어졌으며 음성의 시작점과 끝난후에 적어도 약 50ms 이상의 묵음부분을 가지고 있다. 그리고 20 kHz로 샘플링(sampling) 되었으며 한 샘플(sample)당 12비트(bit)로 양자화 되었다. 그외에 75 Hz-8 kHz의 대역통과 필터(band pass filter)로 필터링(filtering) 되어졌다.

다음은 실험에 사용한 자음과 모음이다.

자음 : ㄱ, ㄲ, ㅋ, ㄴ, ㄷ, ㄸ, ㅌ, ㄹ, ㅁ, ㅂ, ㅃ, ㅍ, ㅅ, ㅆ, ㅈ, ㅉ, ㅊ, ㅎ.

모음 : ㅏ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ, ㅐ.

2. 인식 방법

단일 프레임 모음 인식 방법[6]을 이용한 화자 종속(speaker dependent) 인식실험을 수행하였다. 먼저 두 사람의 첫번째 발음한 126개의 각 음성 데이타에 20ms를 한 프레임으로 하는 사각형 창(rectangular window)을 씌워 중첩없이 이동시켜 양쪽으로 6 dB까지 떨어지는 프레임들을 모음 부분의 안정된 영역으로 간주하여 이들을 추출한다. 이때 6 dB의 값은 실험적으로 구하였다. 그리고 한 모음당 18개의 선행 자음이 오므로 동일한 모음을 가진 18개의 음성 데이타에서 추출한 모음 프레임들을 모두 모아 그 모음에 대한 LPC 켑스트럼과 프랙탈 차원의 코드북(codebook)을 만들기 위한 학습 데이타로서 사용하였다. 이때

각 모음의 학습 데이타들은 250에서 600프레임 사이의 크기를 가지고 있다. LPC 켑스트럼은 각 모음의 학습 데이타를 먼저 preemphasis 하고 각 프레임당 헤밍창(Hamming window)을 씌워 10ms씩 중첩시키면서 LPC 계수를 먼저 구하여 LPC 켑스트럼을 구하였다. LPC 계수를 구할 때 자기상관 함수(auto-correlation function)를 이용하는 Durbin 알고리즘[9]을 사용하였다. 프랙탈 차원은 preemphasis와 헤밍창을 사용하지 않고 단지 각 프레임을 10ms씩 중첩시키면서 구하였다. 이때 프랙탈 차원은 앞에서 언급했듯이 $\varepsilon_0=1$인 10, 20, 30LFD를 사용하였다. 그리고 각각의 코드북은 LBG 알고리즘[10]을 이용해 구성하였다.

다음으로는 만들어진 코드북을 사용하여 인식실험을 수행하는 과정을 설명한다. 각 사람의 두번째 발음한 126개의 음성 데이타(코드북을 만드는데 사용하지 않은 데이타)에서 가장 높은 에너지를 갖는 한 프레임을 추출하여 역시 LPC 켑스트럼을 구할 때는 preemphasis 하고 헤밍창을 씌워 LPC 켑스트럼을 구하고 프랙탈 차원은 앞에서와 같이 preemphasis와 헤밍창을 사용하지 않고 $\varepsilon_0=1$인 10, 20, 30LFD를 구하였다. 다음으로 126개 각각에서 구한 LPC 켑스트럼과 프랙탈 차원을 각 모음(7개 모음)의 LPC 켑스트럼과 프랙탈 차원 코드북과 비교하여 LPC 켑스트럼과 프랙탈 차원 각각의 거리를 구한다. 그리고 각 거리에 서로 다른 가중치(weight)를 주어 더한 전체거리가 가장 작은값을 가지는 코드북의 모음을 인식된 모음으로 결정한다. 본 논문에서 LPC 켑스트럼의 차수 p와 각각의 코드북 크기는 실험적으로 결정하였다(제 4장 참조). 여기서 LPC 켑스트럼과 프랙탈 차원, 그리고 전체거리는 다음과 같다.

(i) 프레임 A와 B의 LPC 켑스트럼을 각각 $C_a(i)$, $C_b(i)$ $(1\le i\le p)$라고 할때 A와 B의 LPC 켑스트럼 거리 $D_{ceps}(A, B)$는 다음과 같이 주어진다.

$$D_{ceps}(A, B)=\sum_{i=1}^{p}[C_a(i)-C_b(i)]^2. \tag{9}$$

(ii) 프레임 A와 B의 프랙탈 차원을 F_a, F_b라 할때 A와 B의 프랙탈 차원거리 $D_{frac}(A, B)$는 다음과 같이 주어진다.

$$D_{frac}(A, B)=[F_a-F_b]^2. \tag{10}$$

(iii) $D_{ceps}(A, B)$의 평균값을 $E[D_{ceps}(\cdot,\cdot)]$라 하고 $D_{frac}(A, B)$의 평균값을 $E[D_{frac}(\cdot,\cdot)]$라 할때 프

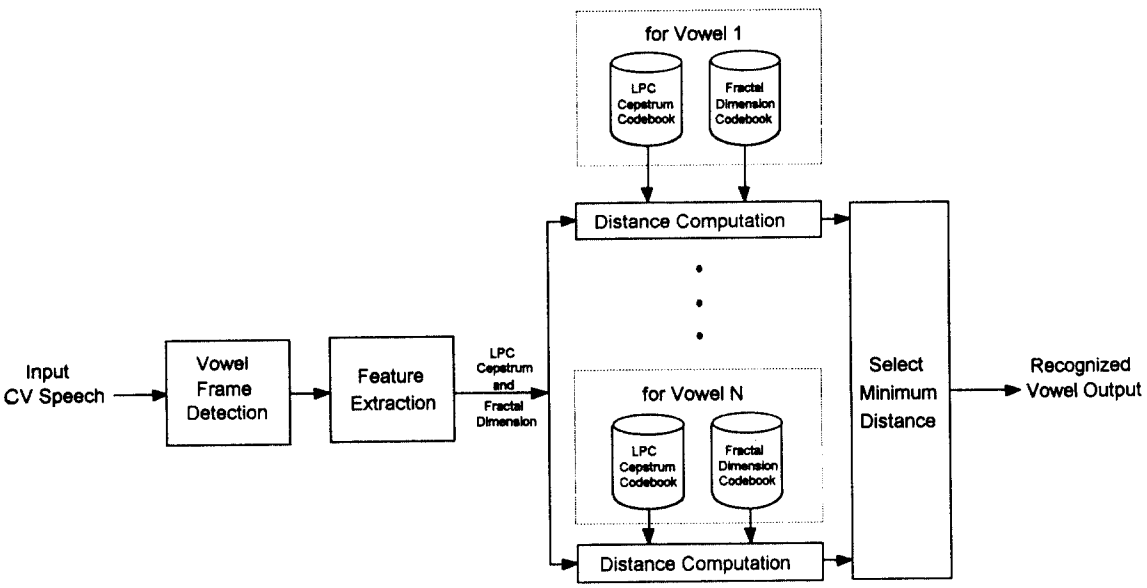

그림 4. LPC켑스트럼과 프랙탈 차원을 이용한 모음 인식 시스템
Fig 4. Vowel recognition system using the LPC cepstra and the fractal dimension

레임 A와 B에 서로 다른 비중을 둔 전체 거리 $D_{total}(A, B)$를 다음과 같이 정의한다.

$$D_{total}(A, B) = \alpha \cdot \frac{D_{ceps}(A, B)}{E[D_{ceps}(\cdot, \cdot)]} + (1-\alpha) \cdot \frac{D_{frac}(A, B)}{E[D_{frac}(\cdot, \cdot)]} \quad (11)$$

여기서 α가 1일 때는 LPC 켑스트럼 거리만 사용한 경우이며, α가 0일 때는 프랙탈 차원의 거리만 사용한 경우이다. LPC 켑스트럼 거리 및 프랙탈 차원의 거리에 대한 평균값은 실험용 음성 데이타로 부터 추정하였다.

실험에서 각 α에 따른 인식결과를 조사하여 가장 높은 인식율을 가지는 α값을 구한다. 지금까지 설명한 인식 시스템을 그림 4에 나타내었다.

Ⅳ. 실험 및 결과의 검토

먼저 LPC 켑스트럼과 프랙탈 차원 각각을 사용할 때 가장 좋은 인식 결과를 나타내는 조건을 구하였다. LPC 켑스트럼을 사용할때 코드북 크기와 LPC 켑스트럼 차수를 각각 1에서 32까지, 16차에서 26차까지 바꾸어 실험하였다. 이때 두 사람 모두 코드북 크기가 4, LPC 켑스트럼 차수가 22일때 가장 좋은 인식 결과를 보였다. 표 1에 LPC 켑스트럼만을 사용했을 경우에 대한 두 사람의 평균 오인식율 결과를 나타내었다. 그리고 프랙탈 차원도 10, 20, 30LFD 각각에 대해 코드북 크기를 1에서 32까지 변화 시키면서 실험하였다. 이때도 두 사람 모두 10, 20, 30LFD의 전 경우에 대해 코드북 크기를 1로 하였을때 가장 좋은 결과를 보였으며 그 중 20LFD를 사용했을때가 가장 좋은 성능을 보였다. 표 2에 프랙탈 차원만을 사용했을 경우에 대한 두 사람의 평균 오인식율을 나타내었다. 물론 표에서 알 수 있는 바와 같이, 프랙탈차원 한 가지만을 음성인식 특징 파라메터로 사용했을 경우의 인식성능은 매우 저조하였다.

그 다음으로 식 (11)에 의거하여 LPC 켑스트럼과 프랙탈 차원을 같이 사용하여 인식 실험을 하였으며, 이 결과를 표 3에 나타내었다. 여기서 α가 1일때는 LPC 켑스트럼만을 사용한 경우이며, 0일때는 프랙탈 차원만을 사용한 경우이다. 표에서 보는 바와 같이 두 사람에 모두에 대해 LPC 켑스트럼과 프랙탈 차원을 함께 사용함으로써 오인식율의 감소를 보였으며, LPC 켑스트럼만을 사용했을 때의 평균 오인식율 5.6%에 비해 프랙탈 차원을 함께 사용함으로써 평균 3.2%의 오인식율을 얻어, 오인식 되는 경우가 40%이상 감소됨을 알 수 있다. 또한, α가 0.9와 0.5 사이의 비교적 넓은 범위에 대해 두 사람 모두 인식율이 증가 하였는데, 이는 α의 변화에 따른 인식율의 민감도(sensitivity)가 적어 본 논문에서 제안한 인식 방식의 유용성을 보여준다. 그리고 평균적으로는 α가 0.5 일때 가장 높은 인식율을 보였는데 이는 식 (11)에서 각각의 거리들에 대해 평균값으로 정규화 시킨 것이 합리적임을 시사한다.

이상으로서 프랙탈 차원은 음성 인식에 있어서 하나의 좋은 파라메터가 될 수 있음을 알수 있었다. 그리고 참고적으로 일반적인 화자종속 단어 인식실험

표 1. 코드북 크기와 LP 켑스트럼 차수에 따른 LPC 켑스트럼을 이용한 모음 오인식율.(두 사람의 평균)

Table 1. Vowel recognition error rate using LPC cepstrum according to various codebook sizes and LPC cepstrum orders.(average of two speakers)

LPC 켑스트럼 차수 / 코드북 크기	16	18	20	22	24	26
1	14.7%	12.3%	13.5%	12.7%	13.5%	13.5%
2	14.3%	12.3%	13.5%	11.9%	11.5%	9.9%
4	11.1%	12.8%	9.9%	5.6%	8.7%	8.3%
8	16.3%	14.3%	12.3%	9.1%	9.5%	11.5%
16	12.7%	11.5%	8.7%	7.5%	7.1%	8.7%
32	9.9%	10.7%	8.7%	7.9%	7.5%	7.5%

표 2. 코드북 크기와 mLFD에 따른 프랙탈 차원을 이용한 모음 오인식율의 변화.(두 사람의 평균)

Table 2. Vowel recognition error rate using fractal dimension according to various codebook sizes and mLFD.(average of two speakers)

mLFD / 코드북 크기	10LFD	20LFD	30LFD
1	44.0%	35.3%	44.4%
2	47.6%	56.0%	51.6%
4	51.6%	51.6%	50.0%
8	58.3%	44.8%	57.9%
16	60.7%	49.2%	57.1%
32	52.0%	45.6%	54.8%

표 3. α의 변화에 따른 모음 오인식율.(LPC 켑스트럼 코드북 크기=4, 20LFD 코드북 크기=1, LPC 켑스트럼 차수=22)

Table 3. Vowel recognition error rate as a function of α.(LPC cepstrum codebook size=4, 20LFD codebook size=1, cepstrum order=22).

사람 / α	HHI	PCK	전체
1.0	4.8%	6.4%	5.6%
0.9	4.0%	5.6%	4.8%
0.8	4.0%	4.8%	4.4%
0.7	4.8%	4.8%	4.8%
0.6	4.0%	3.2%	3.6%
0.5	3.2%	3.2%	3.2%
0.4	7.1%	3.2%	5.2%
0.3	7.9%	5.6%	6.8%
0.2	11.9%	7.9%	9.9%
0.1	22.2%	11.9%	17.1%
0.0	40.5%	30.2%	35.3%

인식율에 비해 본 논문에서의 모음 인식율이 비교적 낮은데 이는 모든 초성자음 환경에서의 모음인식, 즉 문맥 독립(context independent)인식 실험에 기인한 것으로 판단된다.

V. 결 론

본 논문에서는 형태학적 커버링 방법으로 구한 Minkowski-Bouligand 차원을 프랙탈 차원으로 사용해서 한국어 모음의 인식 실험을 수행하였다. 두 사람의 화자에 대해 LPC 켑스트럼과 프랙탈 차원을 같이 사용한 화자 종속 인식 실험을 한 결과, 넓은 범위의 LPC 켑스트럼 거리 및 프랙탈 차원 거리의 조합에 대해 인식율이 증가하였다. 그리고 LPC 켑스트럼과 프랙탈 차원에 대한 정규화된 가중치가 동일한 경우 가장 우수한 인식 성능을 나타냈으며, 두 사람의 화자에 대한 평균 모음 오인식율이 5.6%에서 3.2%로 감소되었다. 이로써 프랙탈 차원은 음성의 인식에 있어서 하나의 유용한 특징 파라메타가 될 수 있음을 확인하였다. 앞으로 보다 많은 사람들의 음성에 대한 인식 실험 및 화자 독립 인식 실험으로 확장이 이루어져야 할 것으로 보이며, 프랙탈 차원의 통계적인 모델링을 이용하여 Hidden Markov Model(HMM)에 의한 단어인식 실험에도 적용할 계획이다.

감사의 글

본 논문에서 실험에 사용한 음성 데이타는 한국전자통신연구소의 자동통역 연구실에서 구성하여 제공한 것이며, 이 지면을 빌어 음성 데이타를 사용하도록 해주신 자동통역연구실 관계자 여러분께 감사드립니다.

참 고 문 헌

1. B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, New York, 1983.
2. M. Barnsley, *Fractals Everywhere*, Acadmic Press, Inc., New York, 1988.
3. P. Maragos, "Fractal Aspects of Speech Signals : Dimension and Interpolation," in *Proc. IEEE ICASSP-91*, Toronto, Canada, pp. 417-420, May 1991.
4. C. Pickover and A. Khorasani, "Fractal Characterization of Speech Waveform Graphs," *Comput. & Graphics*, Vol. 10, No. 1, pp.51-61, 1986.

5. P. Maragos and F. K. Sun, "Mesasuring the Fractal Dimension of Signals : Morphological Covers and Iterative Optimization," *IEEE Trans. on Signal Processing*, Vol. 41, No. 1, pp. 108-121, January 1993.
6. J. Feder, *Fractals*, Plenum Press, New York, 1988.
7. 이용주, 임연자, 한남용, 최준혁, 정유현, "ETRI의 음성 및 텍스트 데이타 베이스의 구축 현황," *제1회 ETRI 음성, 언어 및 음향정보처리 워크샵*, pp. 161-177, 1993.
8. L. R. Rabiner and F. K. Soong, "Single-Frame Vowel Recognition Using Vector Quantization With Several Distance Measures," *AT&T Technical Journal* Vol. 64, No. 10, pp. 2319-2330, December 1985.
9. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
10. Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," IEEE Trans. on Communications, Vol. Com-28, No. 1, pp. 84-95, January 1980.

崔 哲 榮(Chul Young Choi) 정회원
1967년 10월 12일생
1992년 2월 : 부산대학교 전자공학과 졸업(공학사)
1994년 2월 : 부산대학교 대학원 전자공학과 졸업(공학석사)
1994년 1월 ~ 현재 : 금성 일렉트론 근무
※주관심분야 : 음성인식, 영상처리, 프랙탈 등

金 炯 淳(Hyung Soon Kim) 정회원
1960년 8월 21일생
1983년 2월 : 서울대학교 전자공학과 졸업(공학사)
1984년 2월 : 한국과학기술원 전기 및 전자공학과 석사과정(박사과정 조기진학)
1989년 2월 : 한국과학기술원 전기 및 전자공학과 졸업(공학박사)
1987년 1월 ~ 1992년 6월 : 디지콤 정보통신연구소 연구부장
1992년 7월 ~ 현재 : 부산대학교 전자공학과 전임강사
※주관심분야 : 음성인식, 음성합성, 디지탈 통신 및 신호처리

金 在 浩(Jae Ho Kim) 정회원
1957년 3월 23일생
1980년 2월 : 부산대학교 전기기계공학과 졸업(공학사)
1982년 2월 : 한국과학기술원 산업전자공학과 졸업(공학석사)
1990년 2월 : 한국과학기술원 전기 및 전자공학과 졸업(공학박사)
1988년 8월 ~ 1992년 2월 : 삼성전자 통신연구소 화상통신연구실 수석연구원
1992년 3월 ~ 1993년 1월 : 삼성전자 자문교수
1992년 3월 ~ 현재 : 부산대학교 전자공학과 조교수
※주관심분야 : 영상처리, 디지탈 신호처리 VLSI 설계, 문자인식 등

孫 慶 植(Kyung Sik Son) 정회원
1950년 3월 25일생
1973년 2월 : 부산대학교 전자공학과 졸업(공학사)
1977년 8월 : 부산대학교 대학원 전자공학과 졸업(공학석사)
1979년 ~ 1982년 : 부산대학교 전자공학과 전임강사
1985년 10월 : 부산대학교 전자공학과 조교수
1991년 8월 : 경북대학교 대학원 전자공학과 졸업(공학박사)
1991년 10월 ~ 현재 : 부산대학교 전자공학과 부교수
※주관심분야 : 디지탈 신호처리, 신경회로망 등